

On Estimation and Optimization of Probability *

Xinjia Chen

February, 2008

Abstract

In this paper, we develop a general approach for probabilistic estimation and optimization. An explicit formula is derived for controlling the reliability of probabilistic estimation based on a mixed criterion of absolute and relative errors. By employing the Chernoff bound and the concept of sampling, the minimization of a probabilistic function is transformed into an optimization problem amenable for gradient descent algorithms.

1 Estimation of Probability

It is a ubiquitous problem to estimate the probability of an event. Such probability can be interpreted as the expectation, $\mathbb{E}[X]$, of a Bernoulli random variable X . More generally, if X is a random variable bounded in interval $[0, 1]$ with mean $\mathbb{E}[X] = \mu \in (0, 1)$, we can draw n i.i.d. samples X_1, \dots, X_n of X and estimate μ as $\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n}$. Since $\hat{\mu}$ is of random nature, it is crucial to control the statistical error. For this purpose, we have

Theorem 1 *Let $\varepsilon_a \in (0, 1)$ and $\varepsilon_r \in (0, 1)$ be real numbers such that $\frac{\varepsilon_a}{\varepsilon_r} + \varepsilon_a \leq \frac{1}{2}$. Let $\delta \in (0, 1)$. Let X_1, \dots, X_n be i.i.d. random variables defined in probability space $(\Omega, \mathcal{F}, \Pr)$ such that $0 \leq X_i \leq 1$ and $\mathbb{E}[X_i] = \mu \in (0, 1)$ for $i = 1 \dots, n$. Let $\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n}$. Then,*

$$\Pr \left\{ |\hat{\mu} - \mu| < \varepsilon_a \text{ or } \left| \frac{\hat{\mu} - \mu}{\mu} \right| < \varepsilon_r \right\} > 1 - \delta \quad (1)$$

provided that

$$n > \frac{\varepsilon_r \ln \frac{2}{\delta}}{(\varepsilon_a + \varepsilon_a \varepsilon_r) \ln(1 + \varepsilon_r) + (\varepsilon_r - \varepsilon_a - \varepsilon_a \varepsilon_r) \ln \left(1 - \frac{\varepsilon_a \varepsilon_r}{\varepsilon_r - \varepsilon_a} \right)}. \quad (2)$$

It should be noted that conventional methods for determining sample sizes are based on normal approximation, see [4] and the references therein. In contrast, Theorem 1 offers a rigorous

*The author is currently with Department of Electrical Engineering, Louisiana State University at Baton Rouge, LA 70803, USA, and Department of Electrical Engineering, Southern University and A&M College, Baton Rouge, LA 70813, USA; Email: chenxinjia@gmail.com

method for determining sample sizes. In the special case that X is a Bernoulli random variable, a numerical approach has been developed by Chen [2] which permits exact computation of the minimum sample size.

2 Optimization of Probability

In many applications, it is desirable to find a vector of real numbers θ to minimize a probability, $p(\theta)$, which can be expressed as

$$p(\theta) = \Pr\{Y(\theta, \mathbf{\Delta}) \leq 0\},$$

where $Y(\theta, \mathbf{\Delta})$ is piece-wise continuous with respect to θ and $\mathbf{\Delta}$ is a random vector. If we define

$$\mu(\lambda, \theta) = \mathbb{E}[e^{-\lambda Y(\theta, \mathbf{\Delta})}],$$

then, applying Chernoff bound [3], we have

$$p(\theta) \leq \inf_{\lambda > 0} \mu(\lambda, \theta).$$

This indicates that we can make $p(\theta)$ small by making $\mu(\lambda, \theta)$ small. Hence, we shall attempt to minimize $\mu(\lambda, \theta)$ with respect to $\lambda > 0$ and θ .

To make the new objective function $\mu(\lambda, \theta)$ more tractable, we take a sampling approach. Specifically, we obtain n i.i.d. samples $\mathbf{\Delta}_1, \dots, \mathbf{\Delta}_n$ of $\mathbf{\Delta}$ and approximate $\mu(\lambda, \theta)$ as

$$g(\lambda, \theta) = \frac{\sum_{i=1}^n e^{-\lambda Y(\theta, \mathbf{\Delta}_i)}}{n}.$$

A critical step is the determination of sample size n so that $g(\lambda, \theta)$ is sufficiently close to $\mu(\lambda, \theta)$. Since $0 < e^{-\lambda Y(\theta, \mathbf{\Delta})} < 1$, an appropriate value of n can be computed based on (2) of Theorem 1.

Finally, we have transformed the problem of minimizing the probability function $p(\theta)$ as the problem of minimizing a piece-wise continuous function $g(\lambda, \theta)$. Since $g(\lambda, \theta)$ is a more smooth function, we can bring all the power of nonlinear programming to solve the problem. An extremely useful tool is the *gradient descent algorithm*, see, e.g. [1] and the references therein.

3 Proof of Theorem 1

To prove the theorem, we shall introduce function

$$g(\varepsilon, \mu) = (\mu + \varepsilon) \ln \frac{\mu}{\mu + \varepsilon} + (1 - \mu - \varepsilon) \ln \frac{1 - \mu}{1 - \mu - \varepsilon}$$

where $0 < \varepsilon < 1 - \mu$. We need some preliminary results.

The following lemma is due to Hoeffding [5].

Lemma 1

$$\Pr\{\hat{\mu} \geq \mu + \varepsilon\} \leq \exp(n g(\varepsilon, \mu)) \quad \text{for } 0 < \varepsilon < 1 - \mu < 1,$$

$$\Pr\{\hat{\mu} \leq \mu - \varepsilon\} \leq \exp(n g(-\varepsilon, \mu)) \quad \text{for } 0 < \varepsilon < \mu < 1.$$

Lemma 2 Let $0 < \varepsilon < \frac{1}{2}$. Then, $g(\varepsilon, \mu)$ is monotonically increasing with respect to $\mu \in (0, \frac{1}{2} - \varepsilon)$ and monotonically decreasing with respect to $\mu \in (\frac{1}{2}, 1 - \varepsilon)$. Similarly, $g(-\varepsilon, \mu)$ is monotonically increasing with respect to $\mu \in (\varepsilon, \frac{1}{2})$ and monotonically decreasing with respect to $\mu \in (\frac{1}{2} + \varepsilon, 1)$.

Proof. Tedious computation shows that

$$\frac{\partial g(\varepsilon, \mu)}{\partial \mu} = \ln \frac{\mu(1 - \mu - \varepsilon)}{(\mu + \varepsilon)(1 - \mu)} + \frac{\varepsilon}{\mu} + \frac{\varepsilon}{1 - \mu}$$

and

$$\frac{\partial^2 g(\varepsilon, \mu)}{\partial \mu^2} = -\frac{\varepsilon^2}{\mu^2(\mu + \varepsilon)} - \frac{\varepsilon^2}{(1 - \mu)^2(1 - \mu - \varepsilon)} < 0$$

for $0 < \varepsilon < 1 - \mu < 1$. Note that

$$\frac{\partial g(\varepsilon, \mu)}{\partial \mu} \Big|_{\mu=\frac{1}{2}} = \ln \frac{1 - 2\varepsilon}{1 + 2\varepsilon} + \varepsilon < 0$$

because

$$\frac{d \left[\ln \frac{1-2\varepsilon}{1+2\varepsilon} + \varepsilon \right]}{d\varepsilon} = -\frac{4}{1 - 4\varepsilon^2} < 0.$$

Moreover,

$$\frac{\partial g(\varepsilon, \mu)}{\partial \mu} \Big|_{\mu=\frac{1}{2}-\varepsilon} = \ln \frac{1 - 2\varepsilon}{1 + 2\varepsilon} + \frac{4\varepsilon}{1 - 4\varepsilon^2} > 0$$

because

$$\frac{d \left[\ln \frac{1-2\varepsilon}{1+2\varepsilon} + \frac{4\varepsilon}{1-4\varepsilon^2} \right]}{d\varepsilon} = \frac{32\varepsilon^2}{(1 - \varepsilon^2)^2} > 0.$$

Similarly,

$$\frac{\partial g(-\varepsilon, \mu)}{\partial \mu} = \ln \frac{\mu(1 - \mu + \varepsilon)}{(\mu - \varepsilon)(1 - \mu)} - \frac{\varepsilon}{\mu} - \frac{\varepsilon}{1 - \mu}$$

and

$$\frac{\partial^2 g(-\varepsilon, \mu)}{\partial \mu^2} = -\frac{\varepsilon^2}{\mu^2(\mu - \varepsilon)} - \frac{\varepsilon^2}{(1 - \mu)^2(1 - \mu + \varepsilon)} < 0$$

for $0 < \varepsilon < \mu < 1$. Hence,

$$\frac{\partial g(-\varepsilon, \mu)}{\partial \mu} \Big|_{\mu=\frac{1}{2}} = \ln \frac{1 + 2\varepsilon}{1 - 2\varepsilon} - \varepsilon > 0$$

because

$$\frac{d \left[\ln \frac{1+2\varepsilon}{1-2\varepsilon} - \varepsilon \right]}{d\varepsilon} = \frac{4}{1 - 4\varepsilon^2} > 0;$$

and

$$\frac{\partial g(-\varepsilon, \mu)}{\partial \mu} \Big|_{\mu=\frac{1}{2}+\varepsilon} = \ln \frac{1+2\varepsilon}{1-2\varepsilon} - \frac{4\varepsilon}{1-4\varepsilon^2} < 0$$

as a result of

$$\frac{d \left[\ln \frac{1+2\varepsilon}{1-2\varepsilon} - \frac{4\varepsilon}{1-4\varepsilon^2} \right]}{d\varepsilon} = -\frac{32\varepsilon^2}{(1-\varepsilon^2)^2} < 0.$$

Since $\frac{\partial g(\varepsilon, \mu)}{\partial \mu} \Big|_{\mu=\frac{1}{2}} < 0$, $\frac{\partial g(\varepsilon, \mu)}{\partial \mu} \Big|_{\mu=\frac{1}{2}-\varepsilon} > 0$ and $g(\varepsilon, \mu)$ is concave with respect to μ , it must be true that $g(\varepsilon, \mu)$ is monotonically increasing with respect to $\mu \in (0, \frac{1}{2} - \varepsilon)$ and monotonically decreasing with respect to $\mu \in (\frac{1}{2}, 1 - \varepsilon)$. Since $\frac{\partial g(-\varepsilon, \mu)}{\partial \mu} \Big|_{\mu=\frac{1}{2}} > 0$, $\frac{\partial g(-\varepsilon, \mu)}{\partial \mu} \Big|_{\mu=\frac{1}{2}+\varepsilon} < 0$ and $g(-\varepsilon, \mu)$ is concave with respect to μ , it must be true that $g(-\varepsilon, \mu)$ is monotonically increasing with respect to $\mu \in (\varepsilon, \frac{1}{2})$ and monotonically decreasing with respect to $\mu \in (\frac{1}{2} + \varepsilon, 1)$. \square

Lemma 3 Let $0 < \varepsilon < \frac{1}{2}$. Then,

$$\begin{aligned} g(\varepsilon, \mu) &> g(-\varepsilon, \mu) & \forall \mu \in \left(\varepsilon, \frac{1}{2} \right], \\ g(\varepsilon, \mu) &< g(-\varepsilon, \mu) & \forall \mu \in \left(\frac{1}{2}, 1 - \varepsilon \right). \end{aligned}$$

Proof. It can be shown that

$$\frac{\partial [g(\varepsilon, \mu) - g(-\varepsilon, \mu)]}{\partial \varepsilon} = \ln \left[1 + \frac{\varepsilon^2(1-2\mu)}{(\mu^2 - \varepsilon^2)(1-\mu)^2} \right]$$

for $0 < \varepsilon < \min(\mu, 1 - \mu)$. Note that

$$\frac{\varepsilon^2(1-2\mu)}{(\mu^2 - \varepsilon^2)(1-\mu)^2} > 0 \quad \text{for } \varepsilon < \mu < \frac{1}{2}$$

and

$$\frac{\varepsilon^2(1-2\mu)}{(\mu^2 - \varepsilon^2)(1-\mu)^2} < 0 \quad \text{for } \varepsilon < \frac{1}{2} < \mu < 1 - \varepsilon.$$

Therefore,

$$\frac{\partial [g(\varepsilon, \mu) - g(-\varepsilon, \mu)]}{\partial \varepsilon} > 0 \quad \text{for } \varepsilon < \mu < \frac{1}{2}$$

and

$$\frac{\partial [g(\varepsilon, \mu) - g(-\varepsilon, \mu)]}{\partial \varepsilon} < 0 \quad \text{for } \varepsilon < \frac{1}{2} < \mu < 1 - \varepsilon.$$

So, we can complete the proof of the lemma by observing the sign of the partial derivative $\frac{\partial [g(\varepsilon, \mu) - g(-\varepsilon, \mu)]}{\partial \varepsilon}$ and the fact that $g(\varepsilon, \mu) - g(-\varepsilon, \mu) = 0$ for $\varepsilon = 0$. \square

Lemma 4 Let $0 < \varepsilon < 1$. Then, $g(\varepsilon\mu, \mu)$ is monotonically decreasing with respect to $\mu \in \left(0, \frac{1}{1+\varepsilon}\right)$. Similarly, $g(-\varepsilon\mu, \mu)$ is monotonically decreasing with respect to $\mu \in (0, 1)$.

Proof. Note that

$$\frac{\partial g(\varepsilon\mu, \mu)}{\partial \mu} = (1 + \varepsilon) \ln \frac{1 - (1 + \varepsilon)\mu}{1 - \mu} - (1 + \varepsilon) \ln(1 + \varepsilon) + \frac{\varepsilon}{1 - \mu}$$

and

$$\frac{\partial^2 g(\varepsilon\mu, \mu)}{\partial \mu^2} = -\frac{\varepsilon^2}{(1 - \mu)^2[1 - (1 + \varepsilon)\mu]} < 0$$

for any $\mu \in \left(0, \frac{1}{1 + \varepsilon}\right)$.

Since $\frac{\partial g(\varepsilon\mu, \mu)}{\partial \mu}|_{\mu=0} = \varepsilon - (1 + \varepsilon) \ln(1 + \varepsilon) < 0$, we have

$$\frac{\partial g(\varepsilon\mu, \mu)}{\partial \mu} < 0, \quad \forall \mu \in \left(0, \frac{1}{1 + \varepsilon}\right)$$

and it follows that $g(\varepsilon\mu, \mu)$ is monotonically decreasing with respect to $\mu \in \left(0, \frac{1}{1 + \varepsilon}\right)$.

Similarly, since

$$\frac{\partial g(-\varepsilon\mu, \mu)}{\partial \mu}|_{\mu=0} = -\varepsilon - (1 - \varepsilon) \ln(1 - \varepsilon) < 0$$

and

$$\frac{\partial^2 g(-\varepsilon\mu, \mu)}{\partial \mu^2} = -\frac{\varepsilon^2}{(1 - \mu)^2[1 - (1 - \varepsilon)\mu]} < 0, \quad \forall \mu \in (0, 1)$$

we have

$$\frac{\partial g(-\varepsilon\mu, \mu)}{\partial \mu} < 0, \quad \forall \mu \in (0, 1)$$

and, consequently, $g(-\varepsilon\mu, \mu)$ is monotonically decreasing with respect to $\mu \in (0, 1)$.

□

Lemma 5 Suppose $0 < \varepsilon_r < 1$ and $0 < \frac{\varepsilon_a}{\varepsilon_r} + \varepsilon_a \leq \frac{1}{2}$. Then,

$$\Pr\{\hat{\mu} \leq \mu - \varepsilon_a\} \leq \exp\left(n g\left(-\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right) \quad (3)$$

for $0 < \mu \leq \frac{\varepsilon_a}{\varepsilon_r}$.

Proof. We shall show (3) by investigating three cases as follows. In the case of $\mu < \varepsilon_a$, it is clear that

$$\Pr\{\hat{\mu} \leq \mu - \varepsilon_a\} = 0 < \exp\left(n g\left(-\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right).$$

In the case of $\mu = \varepsilon_a$, we have

$$\begin{aligned}
\Pr\{\hat{\mu} \leq \mu - \varepsilon_a\} &= \Pr\{\hat{\mu} = 0\} = \Pr\{X_i = 0, i = 1, \dots, n\} \\
&= \sum_{i=1}^n \Pr\{X_i = 0\} = (\Pr\{X = 0\})^n \\
&= (1 - \Pr\{X \neq 0\})^n \leq (1 - \mathbb{E}[X])^n \\
&= (1 - \mu)^n = (1 - \varepsilon_a)^n \\
&= \lim_{\mu \rightarrow \varepsilon_a} \exp(n g(-\varepsilon_a, \mu)) \\
&< \exp\left(n g\left(-\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right),
\end{aligned}$$

where the last inequality follows from Lemma 2 and the fact that $\varepsilon_a < \frac{\varepsilon_a}{\varepsilon_r} \leq \frac{1}{2} - \varepsilon_a$.

In the case of $\varepsilon_a < \mu \leq \frac{\varepsilon_a}{\varepsilon_r}$, we have

$$\Pr\{\hat{\mu} \leq \mu - \varepsilon_a\} \leq \exp(n g(-\varepsilon_a, \mu)) < \exp\left(n g\left(-\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right),$$

where the first inequality follows from Lemma 1 and the second inequality follows from Lemma 2 and the fact that $\varepsilon_a < \frac{\varepsilon_a}{\varepsilon_r} \leq \frac{1}{2} - \varepsilon_a$. So, (3) is established. \square

Lemma 6 Suppose $0 < \varepsilon_r < 1$ and $0 < \frac{\varepsilon_a}{\varepsilon_r} + \varepsilon_a \leq \frac{1}{2}$. Then,

$$\Pr\{\hat{\mu} \geq (1 + \varepsilon_r)\mu\} \leq \exp\left(n g\left(\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right) \quad (4)$$

for $\frac{\varepsilon_a}{\varepsilon_r} < \mu < 1$.

Proof. We shall show (4) by investigating three cases as follows. In the case of $\mu > \frac{1}{1+\varepsilon_r}$, it is clear that

$$\Pr\{\hat{\mu} \geq (1 + \varepsilon_r)\mu\} = 0 < \exp\left(n g\left(\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right).$$

In the case of $\mu = \frac{1}{1+\varepsilon_r}$, we have

$$\begin{aligned}
\Pr\{\hat{\mu} \geq (1 + \varepsilon_r)\mu\} &= \Pr\{\hat{\mu} = 1\} = \Pr\{X_i = 1, i = 1, \dots, n\} \\
&= \sum_{i=1}^n \Pr\{X_i = 1\} = (\Pr\{X = 1\})^n \\
&\leq \mu^n = \left(\frac{1}{1 + \varepsilon_r}\right)^n \\
&= \lim_{\mu \rightarrow \frac{1}{1+\varepsilon_r}} \exp(n g(\varepsilon_r \mu, \mu)) \\
&< \exp\left(n g\left(\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right),
\end{aligned}$$

where the last inequality follows from Lemma 4 and the fact that $\frac{\varepsilon_a}{\varepsilon_r} \leq \frac{1}{2} \frac{1}{1+\varepsilon_r} < \frac{1}{1+\varepsilon_r}$ as a result of $0 < \frac{\varepsilon_a}{\varepsilon_r} + \varepsilon_a \leq \frac{1}{2}$.

In the case of $\frac{\varepsilon_a}{\varepsilon_r} < \mu < \frac{1}{1+\varepsilon_r}$, we have

$$\Pr\{\hat{\mu} \leq (1 + \varepsilon_r)\mu\} \leq \exp(n g(\varepsilon_r\mu, \mu)) < \exp\left(n g\left(\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right),$$

where the first inequality follows from Lemma 1 and the second inequality follows from Lemma 4. So, (4) is established. \square

We are now in a position to prove the theorem. We shall assume (2) is satisfied and show that (1) is true. It suffices to show that

$$\Pr\{|\hat{\mu} - \mu| \geq \varepsilon_a, |\hat{\mu} - \mu| \geq \varepsilon_r\mu\} < \delta.$$

For $0 < \mu \leq \frac{\varepsilon_a}{\varepsilon_r}$, we have

$$\begin{aligned} \Pr\{|\hat{\mu} - \mu| \geq \varepsilon_a, |\hat{\mu} - \mu| \geq \varepsilon_r\mu\} &= \Pr\{|\hat{\mu} - \mu| \geq \varepsilon_a\} \\ &= \Pr\{\hat{\mu} \geq \mu + \varepsilon_a\} + \Pr\{\hat{\mu} \leq \mu - \varepsilon_a\}. \end{aligned} \quad (5)$$

Noting that $0 < \mu + \varepsilon_a \leq \frac{\varepsilon_a}{\varepsilon_r} + \varepsilon_a \leq \frac{1}{2}$, we have

$$\Pr\{\hat{\mu} \geq \mu + \varepsilon_a\} \leq \exp(n g(\varepsilon_a, \mu)) \leq \exp\left(n g\left(\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right),$$

where the first inequality follows from Lemma 1 and the second inequality follows from Lemma 2. It can be checked that (2) is equivalent to

$$\exp\left(n g\left(\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right) < \frac{\delta}{2}.$$

Therefore,

$$\Pr\{\hat{\mu} \geq \mu + \varepsilon_a\} < \frac{\delta}{2}$$

for $0 < \mu \leq \frac{\varepsilon_a}{\varepsilon_r}$.

On the other hand, since $\varepsilon_a < \frac{\varepsilon_a}{\varepsilon_r} < \frac{1}{2}$, by Lemma 5 and Lemma 3, we have

$$\Pr\{\hat{\mu} \leq \mu - \varepsilon_a\} \leq \exp\left(n g\left(-\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right) \leq \exp\left(n g\left(\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right) < \frac{\delta}{2}$$

for $0 < \mu \leq \frac{\varepsilon_a}{\varepsilon_r}$. Hence, by (5),

$$\Pr\{|\hat{\mu} - \mu| \geq \varepsilon_a, |\hat{\mu} - \mu| \geq \varepsilon_r\mu\} < \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

This proves (1) for $0 < \mu \leq \frac{\varepsilon_a}{\varepsilon_r}$.

For $\frac{\varepsilon_a}{\varepsilon_r} < \mu < 1$, we have

$$\begin{aligned} \Pr\{|\hat{\mu} - \mu| \geq \varepsilon_a, |\hat{\mu} - \mu| \geq \varepsilon_r \mu\} &= \Pr\{|\hat{\mu} - \mu| \geq \varepsilon_r \mu\} \\ &= \Pr\{\hat{\mu} \geq \mu + \varepsilon_r \mu\} + \Pr\{\hat{\mu} \leq \mu - \varepsilon_r \mu\}. \end{aligned}$$

Invoking Lemma 6, we have

$$\Pr\{\hat{\mu} \geq \mu + \varepsilon_r \mu\} \leq \exp\left(n g\left(\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right).$$

On the other hand,

$$\Pr\{\hat{\mu} \leq \mu - \varepsilon_r \mu\} \leq \exp(n g(-\varepsilon_r \mu, \mu)) \leq \exp\left(n g\left(-\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right) \leq \exp\left(n g\left(\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right)$$

where the first inequality follows from Lemma 1, the second inequality follows from Lemma 4, and the last inequality follows from Lemma 3. Hence,

$$\Pr\{|\hat{\mu} - \mu| \geq \varepsilon_a, |\hat{\mu} - \mu| \geq \varepsilon_r \mu\} \leq 2 \exp\left(n g\left(\varepsilon_a, \frac{\varepsilon_a}{\varepsilon_r}\right)\right) < \delta.$$

This proves (1) for $\frac{\varepsilon_a}{\varepsilon_r} < \mu < 1$. The proof of Theorem 1 is thus completed.

References

- [1] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, *Nonlinear Programming – Theory and Algorithms*, Wiley, 1993.
- [2] X. Chen, “Exact computation of minimum sample size for estimation of binomial parameters,” arXiv:0707.2113 [math.ST], July 2007.
- [3] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.
- [4] M. M. Desu and D. Raghavarao, *Sample Size Methodology*, Academic Press, 1990.
- [5] Hoeffding, W. (1963). Probability inequalities for sums of bounded variables. *J. Amer. Statist. Assoc.* **58** 13–29.